

STATISTICA SOCIALE

Dr Paolo Pasetti

Programma del corso

- Prima parte: ANALISI DEI DATI
- Le scale di misurazione
- Organizzare i dati per l'analisi: le matrici di dati
- Analisi monovariata: le misure della tendenza centrale
- Analisi monovariata: Le misure della variabilità

Programma del corso

- Analisi bivariata per variabili categoriali: le tabelle di contingenza
- Analisi bivariata per variabili quantitative: correlazione e regressione
- Calcolo delle probabilità: concetti introduttivi
- Le distribuzioni di probabilità
- Dalla popolazione al campione: le distribuzioni campionarie e gli stimatori

28-29 settembre 2011

Statistica sociale

3

Programma del corso

- Inferenza statistica: intervalli di confidenza
- Inferenza statistica: test di ipotesi
- Analisi elementare delle serie storiche
- Rapporti statistici e numeri indici

28-29 settembre 2011

Statistica sociale

4

Programma del corso

- Seconda parte: LA PROGETTAZIONE DI UN'INDAGINE STATISTICA
- Le tappe di un'indagine campionaria
- Dimensioni della qualità dei dati statistici
- I principali disegni di campionamento
- Gli strumenti di misura: l'intervista e il questionario
- Progettazione del questionario
- Tecniche di somministrazione del questionario

28-29 settembre 2011

Statistica sociale

5

Programma del corso

- Misurare gli atteggiamenti: le tecniche di *scaling*
- Statistica e trappole: gli errori nella ricerca sociale
- Concetti e indicatori: gli indicatori sociali

28-29 settembre 2011

Statistica sociale

6

Testi

- Testi di riferimento
- P.Pasetti, *Statistica del turismo*, Carocci, Roma 2002 (capitoli 1-8);
- D.F.Iezzi, *Statistica per le scienze sociali*, Carocci, Roma 2009.

28-29 settembre 2011

Statistica sociale

7

Testi

- Testi di consultazione
- F.P.Borazzo, P.Perchinunno, *Analisi statistiche con Excel*, Pearson Education Paravia, Milano 2007;
- I.Diamond, J.Jefferies, *Introduzione alla statistica per le scienze sociali (II edizione)*, McGraw-Hill Italia, Milano 2006;
- R.Albano, S.Testa, *Statistica per la ricerca sociale*, Carocci, Roma 2002;
- A.Bosco, *Come si costruisce un questionario*, Carocci, Roma 2003;
- A cura di R.Miceli, *Numeri, dati, trappole*, Carocci, Roma 2004.

28-29 settembre 2011

Statistica sociale

8

ALCUNE DEFINIZIONI

■ **STATISTICA :**

- è una disciplina di carattere **METODOLOGICO** (cioè, **NON È** una scienza sostanziale, come sono invece la fisica, la biologia, l'economia, la geografia, ecc.), basata sull'applicazione di procedimenti di tipo **MATEMATICO**;
- La **METODOLOGIA STATISTICA** si applica tutte le volte che si ha a che fare con un **FENOMENO COLLETTIVO**, cioè con un fenomeno che interessa una **PLURALITÀ** di "oggetti", intesi nel senso più ampio possibile.

28-29 settembre 2011

Statistica sociale

9

ALCUNE DEFINIZIONI

- Tali "oggetti" possono essere persone, oggetti concreti del mondo reale, ma anche eventi, istanti di tempo, ecc. Ad esempio, una **SERIE STORICA** è una successione ordinata di "valori numerici" attribuiti ad una serie di istanti successivi di tempo.
- Nella nostra terminologia, chiameremo questi "oggetti" **UNITÀ STATISTICHE**.
- Non basta, però, poter disporre di un fenomeno collettivo per poter applicare le metodologie statistiche. Occorre, anche, che a ciascuna unità statistica venga assegnato un certo **ATTRIBUTO**, che chiameremo **VARIABLE** (o **CARATTERE**).

28-29 settembre 2011

Statistica sociale

10

La statistica sociale:

- È un insieme di metodologie
- (dal greco: $\mu\epsilon\theta\omicron\delta\omicron\varsigma$ = la via per...) per la raccolta e l'analisi di dati, il cui utilizzo consente di descrivere, valutare e prevedere fenomeni sociali.

La statistica sociale:

- Progetta: pianifica la raccolta dei dati necessari per le ricerche;
- Analizza: illustra, predisponendo opportune sintesi, un insieme di fenomeni collettivi;
- Prevede: l'osservazione e l'analisi di un campione statistico permettono (talvolta) di formulare previsioni.

Il problema della misurazione e le scale di misura

28-29 settembre 2011

Statistica sociale

13

Che cosa significa "misurare"?

- Misurare significa mettere in atto un **PROCESSO** che porta alla produzione di un **DATO STATISTICO**

28-29 settembre 2011

Statistica sociale

14

Che cosa significa "misurare"?

- Secondo il filosofo della scienza KARL POPPER esistono TRE MONDI:
- Il MONDO 1 è il mondo fisico, il mondo delle cose e degli oggetti;
- Il MONDO 2 è il mondo delle idee, dei concetti, dei costrutti teorici, delle formulazioni scientifiche che si riferiscono al MONDO 1;
- Il MONDO 3 è il mondo dei simboli, dei segni e delle regole che "traducono" in termini simbolici e linguistici i concetti del MONDO 2: è il mondo del LINGUAGGIO.

28-29 settembre 2011

Statistica sociale

15

Un esempio

- Al Mondo 1 appartiene lo studente MARIO ROSSI;
- Al Mondo 2 appartiene il concetto: PROFITTO IN MATEMATICA di Mario Rossi;
- Come si traduce, nel Mondo 3, il PROFITTO IN MATEMATICA di Mario Rossi?

28-29 settembre 2011

Statistica sociale

16

Un esempio

- Si traduce con un PUNTEGGIO in una PROVA alla quale è stato sottoposto Mario Rossi; eventualmente si tradurrà poi, alla fine di un percorso, in una media di punteggi;
- Questa operazione, che consiste nel "mettere in relazione" il mondo 1 (il mondo delle cose) con il mondo 2 (il mondo dei concetti) si dice DEFINIZIONE OPERATIVA di un concetto, o anche OPERATIVIZZAZIONE

28-29 settembre 2011

Statistica sociale

17

Riassumendo

- Mondo 1: Mario Rossi;
- Mondo 2: profitto in matematica di Mario Rossi nel primo quadrimestre;
- Mondo 3: un indicatore di sintesi (es.media) dei voti nelle prove di matematica ottenuti da Mario Rossi.

28-29 settembre 2011

Statistica sociale

18

Un altro esempio

- A ogni studente del I anno (**unità statistica**) verrà chiesto:
- Il tipo di scuola di provenienza
- Una valutazione di qualità sulla propria preparazione in vista degli studi universitari
- L'anno in cui è nato/a
- Il voto di maturità (in centesimi)

28-29 settembre 2011

Statistica sociale

19

SUPPONIAMO CHE VOI SIATE STUDENTI ISCRITTI AL PRIMO ANNO

DATO UNO: QUALE SCUOLA SUPERIORE
HA FREQUENTATO PRIMA DI ISCRIVERSI ALL'UNIVERSITÀ?

- ISTITUTO TECNICO COMMERCIALE
- ISTITUTO TECNICO PER GEOMETRI
- ISTITUTO TECNICO INDUSTRIALE
- ALTRO ISTITUTO TECNICO
- ISTITUTO PROFESSIONALE AD INDIRIZZO COMMERCIALE
- ISTITUTO PROFESSIONALE ALBERGHIERO
- ISTITUTO PROFESSIONALE AGRARIO
- ISTITUTO PROFESSIONALE INDUSTRIALE/ARTIGIANALE
- LICEO CLASSICO
- LICEO SCIENTIFICO
- LICEO ARTISTICO
- ISTITUTO MAGISTRALE O ALTRO LICEO
- ALTRO ISTITUTO PROFESSIONALE _____
- ALTRA SCUOLA SUPERIORE _____

28-29 settembre 2011

Statistica sociale

20

DATO DUE: LIVELLO AUTOPERCEPITO DELLA PREPARAZIONE SCOLASTICA

- ADESSO VI CHIEDO DI ESPRIMERE UN'**OPINIONE**.
- LA DOMANDA È:
 - **"CHE VALUTAZIONE DA', IN GENERALE, DELLA SUA PREPARAZIONE SCOLASTICA, IN VISTA DEGLI STUDI UNIVERSITARI?"**
 - **4)** __ Ottima
 - **3)** __ Buona
 - **2)** __ Sufficiente
 - **1)** __ Insufficiente
- chiariremo nel seguito il significato dei "numeri" che abbiamo scritto prima dei quadratini relativi alle risposte

28-29 settembre 2011

Statistica sociale

21

DATO TRE: ANNO DI NASCITA

- Le chiedo adesso di dirmi il suo anno di nascita; si tratta di un dato apparentemente molto banale, che moltissime persone certamente le hanno già chiesto (oltretutto, le ho chiesto solo l'anno, senza chiedere la data completa):
- ANNO DI NASCITA: **19**__ __
- In realtà questo dato non è così banale come sembra; nel seguito vedremo perché.

28-29 settembre 2011

Statistica sociale

22

DATO QUATTRO: **VOTO DI MATURITA'**

- MI PUO' DIRE QUAL E' STATO IL SUO VOTO (ESPRESSO IN CENTESIMI) ALL'ESAME DI MATURITA' ?

■ _____ /100

28-29 settembre 2011

Statistica sociale

23

ABBIAMO COSI' RACCOLTO, SULLA NOSTRA UNITA' STATISTICA, CHE SI CHIAMA:

- I QUATTRO **DATI** CHE SEGUONO:
- SCUOLA _____
- LIVELLO DI PREPARAZIONE SCOLASTICA _____
- ANNO DI NASCITA _____
- VOTO DI MATURITA' _____/100
- Vedremo, nel seguito, che sono tutti **dati** (corrispondono, cioè, alla definizione di "dato" che daremo), ma, soprattutto, che si tratta di quattro "tipi" diversi di dato (SCALE DI MISURA).

28-29 settembre 2011

Statistica sociale

24

Proviamo a ripercorrere le quattro "grandezze" che abbiamo appena rilevato su uno di voi: è evidente che ciascuna di esse rappresenta una **VARIABLE**:

- **SCUOLA DI PROVENIENZA ;**
 - **LIVELLO DI PREPARAZIONE SCOLASTICA ;**
 - **ANNO DI NASCITA ;**
 - **VOTO DI MATURITA' (IN CENTESIMI).**
- Nel momento in cui una variabile (ad esempio, una di quelle appena viste) assume un certo **valore** in corrispondenza di una **UNITA' STATISTICA**, siamo in presenza di un **DATO STATISTICO**.

28-29 settembre 2011

Statistica sociale

25

IL DATO STATISTICO

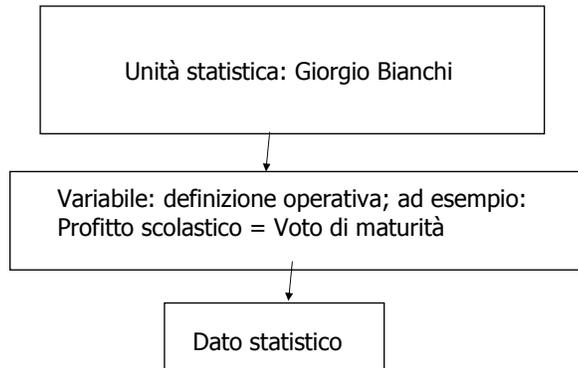
- Così, i quattro **ENUNCIATI**:
 - «lo studente **A.B.** proviene dalla scuola **X**»;
 - «lo studente **A.B.** ha dichiarato il livello di preparazione **Y**»;
 - «lo studente **A.B.** è nato nell'anno **Z**»;
 - «lo studente **A.B.** ha avuto un voto di maturità pari a **W**»
- rappresentano tutti dei **DATI STATISTICI**. Il significato degli ultimi due enunciati come dati statistici verrà chiarito nel seguito.
- Un dato statistico quindi, contrariamente a quanto si ritiene comunemente, non è qualcosa di "oggettivo", ma è anch'esso sottoposto ad un preciso **PROCESSO DI FORMAZIONE**.

28-29 settembre 2011

Statistica sociale

26

Il processo di formazione del dato statistico



28-29 settembre 2011

Statistica sociale

27

Il processo di formazione del dato statistico

- Come si vede, quindi, intervengono nella formazione del dato statistico **DUE LIVELLI**: in entrambi questi livelli sono presenti fattori almeno in parte **CONVENZIONALI**.
- In altre parole, sia per quanto riguarda la scelta dell'**UNITÀ STATISTICA**, che per quanto riguarda la scelta della **VARIABILE DI STUDIO**, il ricercatore deve fissare un **INSIEME DI REGOLE** che permetta di effettuare entrambe tali scelte in modo **NON AMBIGUO**, cioè non più legato alla "definizione" di senso comune.
- Questo insieme di regole è detto, come anticipato sopra, **DEFINIZIONE OPERATIVA** o **OPERATIVIZZAZIONE**.

28-29 settembre 2011

Statistica sociale

28

ALTRA IMPORTANTE DEFINIZIONE :

- **COLLETTIVO** o **POPOLAZIONE:**
 - è l'**INSIEME** delle unità statistiche che sono oggetto di indagine.
- La **STATISTICA DESCRITTIVA** si propone lo scopo principale di fornire indicatori e misure riferite ad un **COLLETTIVO**. Se, invece, non si ha a che fare con un collettivo, ma con un suo sottoinsieme, si parla allora di **CAMPIONE** e, in tal caso, si parla di **STATISTICA INFERENZIALE** o, meglio, di **INFERENZA STATISTICA**.

28-29 settembre 2011

Statistica sociale

29

FUNZIONE DELLA STATISTICA APPLICATA: DAL DATO ALL'INFORMAZIONE

- ***Data is not information***
- ***Information is not knowledge***
- Finora abbiamo visto come costruire un dato statistico. I dati statistici, però, non sono altro che la materia prima, i "mattoni" fondamentali su cui si basa un'indagine statistica. Scopo fondamentale della statistica è, infatti, quello di **FORNIRE INFORMAZIONI RILEVANTI SU UN DETERMINATO FENOMENO**.
- Alla metodologia statistica si rivolgono "domande" su un certo fenomeno, e la statistica risponde fornendo, appunto, **INFORMAZIONI** che siano in grado di cogliere nel miglior modo possibile un certo aspetto di quel fenomeno che viene considerato rilevante.

28-29 settembre 2011

Statistica sociale

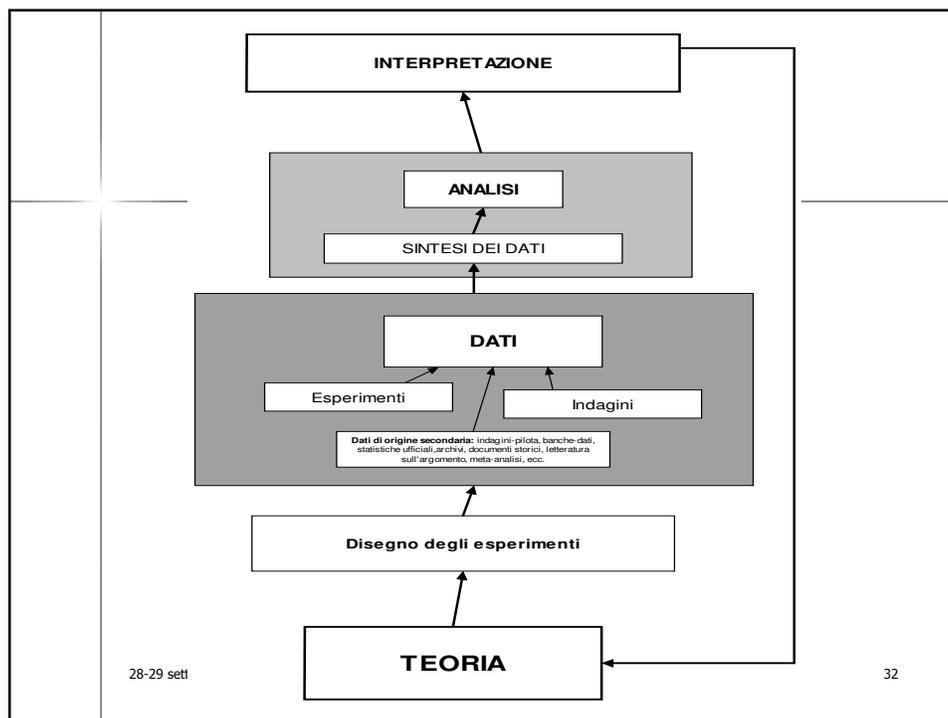
30

- Per produrre informazioni, la statistica deve MANIPOLARE in modo opportuno i dati, fornendo SINTESI SIGNIFICATIVE dei dati stessi. Per usare un concetto formulato da Bertrand Russell, pur basandosi sui dati statistici "individuali", la statistica ha lo scopo di produrre nuovi "dati", che sono però di un "tipo logico" superiore.
- I "dati" di questo "tipo logico" superiore sono appunto le INFORMAZIONI
- Se, ad esempio, ci viene posta la seguente domanda:
"qual è stato, in media, il voto di maturità in una certa scuola di Ferrara? "
- non sarà sufficiente disporre dei singoli voti di maturità relativi ai singoli studenti (DATI), ma si dovrà calcolare il VALORE MEDIO di tutti i voti degli studenti di quella certa scuola (INFORMAZIONE), valore che potrà dare così un'indicazione di sintesi sul voto di maturità di quella scuola e, indirettamente, sul "successo formativo" della scuola stessa.

28-29 settembre 2011

Statistica sociale

31



28-29 sett

32

Tipi di variabile: le scale di misura

- La prima formulazione della teoria delle scale di misurazione è dovuta a STEVENS:
- Articolo dal titolo: "On the Theory of Scales of Measurement", apparso su *Science* nel 1946

28-29 settembre 2011

Statistica sociale

33

Tipi di variabile: le scale di misura

- Ciò che differenzia i diversi tipi di variabile statistica è la **SCALA DI MISURA** o **SCALA DI MISURAZIONE** nella quale sono espressi i dati.
- In termini formali, una SCALA DI MISURAZIONE è rappresentata dalla TERNA di elementi:
 - $\langle SE, f, SN \rangle$
 - dove:
 - **SE** = SISTEMA RELAZIONALE EMPIRICO (Il Mondo 2 di Popper)
 - Esempio: lo studente A è *più bravo* dello studente B
 - **SN** = SISTEMA RELAZIONALE NUMERICO (Il Mondo 3)
 - Esempio: voto (A) = **100** > **74** = voto(B)
- f = funzione che associa ciascun elemento di SE ad uno ed un solo elemento di SN.
- Esempio: f = definizione operativa di "voto di maturità" (con quali criteri e definizioni si costruisce questo voto)

28-29 settembre 2011

Statistica sociale

34

La scala nominale

- Si tratta di un semplice **ATTRIBUTO QUALITATIVO** che viene attribuito alla unità statistica. La variabile "scuola frequentata" è su scala nominale.
- Il tipo di dato statistico che viene generato da una variabile su scala nominale è detto **FREQUENZA** o anche **FREQUENZA ASSOLUTA**.
- La frequenza assoluta è il conteggio di quante unità statistiche assumono una modalità della variabile.

28-29 settembre 2011

Statistica sociale

35

La scala nominale

- Ad esempio, se supponiamo che le matricole siano **200**, possiamo costruire la distribuzione per tipologia di scuola frequentata:

	Freq.	Freq.rel.	%
■ LICEI	50	0,25	25%
■ SCUOLE TECNICHE	70	0,35	35%
■ SCUOLE PROFESSIONALI	80	0,40	40%
■ TOTALE	200	1	100%

- Se la frequenza di una modalità viene rapportata al totale, si ha la **FREQUENZA RELATIVA**.
- La frequenza relativa, moltiplicata per 100, dà il **VALORE PERCENTUALE**.

28-29 settembre 2011

Statistica sociale

36

Relazione ammissibile

- In questa scala di misura, l'unica relazione ammissibile tra due elementi è
- '=' oppure '≠'
- Esempio: se Giorgio e Giovanni vengono da un liceo, mentre Francesca ha frequentato un istituto tecnico, (consideriamo cioè la variabile: **scuola frequentata**), possiamo scrivere:
- GIORGIO = GIOVANNI
- mentre
- GIOVANNI ≠ FRANCESCA

28-29 settembre 2011

Statistica sociale

37

La scala ordinale

- Se, tra i vari attributi considerati, è possibile instaurare un ordine, si parla di scala ordinale;
- Il giudizio sulla propria preparazione scolastica (espresso da un aggettivo: ottima, buona, sufficiente, insufficiente) è un esempio di scala ordinale.

28-29 settembre 2011

Statistica sociale

38

La scala ordinale

- Possiamo costruire la distribuzione delle 200 matricole, a seconda di come hanno risposto al secondo quesito:

	Freq.ass.	Freq.rel.	%
Ottima	20	0,1	10
Buona	60	0,3	30
Sufficiente	100	0,5	50
Insufficiente	20	0,1	10
Totale	200	1	100

Stavolta, le modalità non sono sconnesse, ma hanno un preciso ordine

Relazione ammissibile

- Sulla scala di misura ordinale, tra un'unità e l'altra vi può essere solo una RELAZIONE D'ORDINE.

Relazione ammissibile

- Esempio: sappiamo che
- risposta(GIOVANNI) = 4
- risposta(FRANCESCA) = 3
- risposta(GIORGIO) = 2
- Pertanto, possiamo scrivere:
- GIOVANNI > FRANCESCA
- FRANCESCA > GIORGIO
- Vale la proprietà transitiva:
- GIOVANNI > GIORGIO
- Non si può fare l'operazione di DIFFERENZA tra un elemento e l'altro:
- GIOVANNI – FRANCESCA = ?

28-29 settembre 2011

Statistica sociale

41

La scala ad intervalli

- Le variabili espresse in queste scala sono VARIABILI QUANTITATIVE vere e proprie.
- La SCALA AD INTERVALLI non ammette uno "zero" assoluto, ma soltanto uno "zero" convenzionale. Questo fa sì che in questa scala siano consentite tra le unità solo operazioni di differenza.
- Esempi classici di scala ad intervalli sono:
 - la scala centigrada (o, in modo del tutto equivalente, la scala Fahrenheit), delle temperature, nella quale lo "zero" è puramente convenzionale;
 - la numerazione degli anni nel calendario cristiano, il cui inizio (nascita di Gesù di Nazareth) è puramente convenzionale.

28-29 settembre 2011

Statistica sociale

42

La scala ad intervalli

- Per tornare al nostro esempio:
- Se Giorgio Bianchi è nato nel 1988, il numero "**1988**" appartiene a una scala ad intervalli, perché il suo "punto zero" non è assoluto ma convenzionale

28-29 settembre 2011

Statistica sociale

43

Relazione ammissibile

- In questa scala, tra due unità esiste una relazione di differenza quantitativa vera e propria;
- Tra due unità è possibile calcolare la differenza:
- anno di nascita (GIORGIO) = 1988 (anni dalla nascita di Gesù di Nazareth)
- anno di nascita (GIOVANNI) = 1986 (anni dalla nascita di Gesù di Nazareth)
- differenza di età tra Giovanni e Giorgio =
- = $(1988 - 1986) = 2$ anni

28-29 settembre 2011

Statistica sociale

44

Relazione NON ammissibile

- Nella scala ad intervalli non è ammessa l'operazione di rapporto tra due unità;
- Questo proprio per la natura convenzionale dello "zero" di questa scala;
- Per esempio, non avrebbe alcun senso, infatti, dire che l'anno "2000" è "il doppio" dell'anno "1000"!

28-29 settembre 2011

Statistica sociale

45

La scala a rapporti

- La SCALA A RAPPORTI, che ammette lo zero assoluto, consente sia l'operazione di differenza, sia l'operazione di rapporto tra le diverse unità.
- Esempi classici di scale a rapporti sono:
 - la scala Kelvin delle temperature, che inizia con lo "zero" assoluto;
 - la scala delle distanze lineari, espresse in metri, chilometri, ecc.;
 - la scala del tempo, espresso in minuti, secondi, nanosecondi, ecc.
- Una scala di valutazione sufficientemente fine, che può essere vista come una quantità vera e propria: ad esempio, il voto di maturità, espresso in centesimi.

28-29 settembre 2011

Statistica sociale

46

Relazione ammissibile

- In questa scala, due unità possono rapportarsi tra loro, perché il "punto zero" è assoluto;
- Tra due unità, pertanto, è possibile calcolare il rapporto:
- voto (FRANCESCA) = 90 / 100
- voto (GIOVANNI) = 60 / 100
- $FRANCESCA / GIOVANNI = 90/60 = 3/2 = 1,5$
- Possiamo dire, quindi, che Francesca, in termini empiricamente calcolabili, è stata **1,5 volte** "più brava" di Giovanni.

28-29 settembre 2011

Statistica sociale

47

Per riassumere

Relazione ammissibile tra A e B	QUALITATIVE		QUANTITATIVE	
	NOMINALE	ORDINALE	AD INTERVALLI	A RAPPORTI
$x_A = x_B ; x_A \neq x_B$	*	*	*	*
$x_A > x_B ; x_A < x_B$		*	*	*
$x_A - x_B ; x_B - x_A$			*	*
$x_A / x_B ; x_B / x_A$				*

28-29 settembre 2011

Statistica sociale

48

COME ORGANIZZARE I DATI PER L'ANALISI: LE MATRICI DI DATI

- Il primo fondamentale strumento per cominciare l'analisi dei dati consiste nella RAPPRESENTAZIONE ORGANIZZATA dei dati stessi
- **MATRICE CASO X VARIABLE**
- TORNIAMO PER UN ATTIMO ALL'ESEMPIO VISTO PRIMA

28-29 settembre 2011

Statistica sociale

49

COME ORGANIZZARE I DATI PER L'ANALISI: LE MATRICI DI DATI

VARIABILI INVESTIGATE

- 1) SCUOLA SUPERIORE FREQUENTATA:
SEMPLIFICANDO, LE MODALITA' SONO: **L** (LICEI), **P** (PROF.LI),
T (TECNICI)
- 2) LIVELLO DI PREPARAZIONE (1,2,3,4)
- 3) ANNO DI NASCITA
- 4) VOTO DI MATURITA' (in centesimi)
- COSTRUIAMO LA MATRICE **CASO X VARIABLE**
- I **CASI**, NEL NOSTRO CASO, SONO LE PERSONE SULLE QUALI
ABBIAMO DECISO DI OSSERVARE LE VARIABILI APPENA VISTE

28-29 settembre 2011

Statistica sociale

50

VARIABILI	SCUOLA	PREP	ANNONASC	VOTO
CASI				
Giacomo	L	2	1990	83
Giuseppe	P	3	1989	94
Maria	L	4	1988	100
Laura	L	3	1990	88
Filippo	T	1	1987	75
Federica	T	4	1990	100

28-29 settembre 2011 Statistica sociale 51

■ IN GENERALE, UNA MATRICE **C X V**,
 ASSUME LA FORMA CHE POTETE
 VEDERE NELLA PROSSIMA
 DIAPOSITIVA

■ SI TRATTA DI UNA MATRICE DI
 DIMENSIONI $M \times k$

28-29 settembre 2011 Statistica sociale 52

	V_1	V_2	...	V_j	...	V_k
1						
2						
...						
i				x_{ij}		
..						
N						

28-29 settembre 2011 Statistica sociale 53

ALTRE MATRICI DI DATI

- **MATRICI "ORIGINE-DESTINAZIONE"**
- Queste matrici di dati sono nate nell'ambito degli studi territoriali e sui trasporti;
- un tipico esempio è la **MATRICE DI PENDOLARISMO**
- Nella matrice che segue, i dati per riga si riferiscono alle persone uscite dai singoli comuni, mentre i dati per colonna si riferiscono alle persone entrate nei singoli comuni (IL 21 OTTOBRE 1991).

28-29 settembre 2011 Statistica sociale 54

DESTINAZIONE	CHIES	FARRA	PIEVE	PUOS	TAMBRE	
ORIGINE						
CHIES	-	14	57	41	14	126
FARRA	5	-	73	57	48	183
PIEVE	10	13	-	97	5	125
PUOS	13	24	89	-	15	141
TAMBRE	5	29	36	45	-	115
	33	80	255	240	82	690

Fonte: Censimento della popolazione, 1991

28-29 settembre 2011 Statistica sociale 55

PROPRIETÀ DI QUESTO TIPO DI MATRICI

- sono QUADRATE, cioè di dimensioni $k \times k$
- NON SONO SIMMETRICHE;
- possono contenere, o anche non contenere, elementi sulla diagonale principale (che, nell'esempio, non sono stati considerati deliberatamente);
- SOMMA TOTALI DI RIGA = SOMMA TOTALI DI COLONNA
- Un altro esempio di matrice del tipo "origine/destinazione" è la **MATRICE DI MOBILITÀ INTER-GENERAZIONALE**, che permette di esaminare la classe sociale attuale in relazione alla CLASSE SOCIALE DELLA FAMIGLIA DI ORIGINE (che si può considerare "classe sociale di partenza")

28-29 settembre 2011 Statistica sociale 56

Classe attuale	Borghesia	Classe media impiegatizia	Piccola borghesia Urbana	Picc. borghesia agricola	Classe operaia urbana	Classe operaia agricola	Totale
Classe di origine							
Borghesia	91	104	31	3	36	1	266
Classe media impiegatizia	91	199	40	1	70	2	403
Picc. borghesia Urbana	109	221	286	6	254	5	881
Picc. borghesia Agraria	21	83	97	60	180	22	463
Classe operaia urbana	128	521	320	8	947	19	1943
Classe operaia Agraria	7	42	55	9	144	33	290
Totale	447	1170	829	87	1631	82	4246

Fonte: Ballarino, Cobalti, 2003
(l'indagine si è svolta nel 1997)

28-29 settembre 2011 Statistica sociale 57

- Si noti il fatto che, in questo caso, SONO PRESENTI GLI ELEMENTI SULLA DIAGONALE PRINCIPALE
 - Come ultimo esempio, vediamo una matrice O-D fornita dall'ISTAT
 - Si tratta degli **ARRIVI** di **ITALIANI** in **STRUTTURE RICETTIVE ALBERGHIERE**, secondo la regione di provenienza e la regione di destinazione, nel corso dell'anno 2000
- 28-29 settembre 2011 Statistica sociale 58

DESTINAZIONE	ARRIVI										
	PROVENIENZA	Piemonte	Valle d'Aosta	Lombardia	Trentino-A. Adige	Veneto	Friuli-V. Giulia	Liguria	Emilia-Romagna	Toscana	Umbria
Piemonte	996.422	78.026	217.177	95.189	227.761	39.228	381.081	350.222	207.388	50.491	63.814
Valle d'Aosta	2.392	4.877	11.929	3.045	6.355	1.847	11.005	25.480	6.317	2.082	1.338
Lombardia	43.265	118.678	1.657.424	584.933	646.274	151.005	721.348	1.154.415	575.336	158.363	243.447
Trentino-A. Adige	5.705	4.104	47.845	100.403	122.381	19.057	18.317	151.973	39.808	11.825	19.191
Veneto	21.881	17.921	226.168	464.410	560.175	124.407	96.069	344.467	212.585	78.987	91.009
Friuli V. Giulia	5.895	4.135	67.797	85.602	124.175	87.374	22.618	88.532	51.240	17.067	18.267
Liguria	12.535	50.141	105.544	94.971	89.713	23.426	98.893	110.799	106.822	28.676	18.368
Emilia-Romagna	18.487	33.328	204.623	393.088	240.193	60.805	124.985	763.819	253.549	70.945	135.722
Toscana	15.809	32.587	176.522	190.208	194.118	32.190	89.261	344.477	444.438	66.381	65.842
Umbria	4.592	5.072	39.522	36.344	41.540	9.374	16.217	139.064	54.398	31.241	46.830
Marche	5.738	8.217	71.459	92.954	77.782	12.760	25.937	120.839	72.020	34.660	75.718
Lazio	26.880	35.559	252.980	226.635	335.522	51.592	142.221	434.431	414.352	174.863	170.579
Abruzzo	4.711	3.262	39.621	26.598	51.768	9.906	18.502	103.351	48.379	24.917	41.765
Molise	2.130	1.443	12.037	4.608	10.232	3.073	5.100	33.126	12.427	5.733	6.833
Campania	16.629	15.900	145.904	54.339	156.420	26.213	91.456	291.435	229.358	132.236	70.090
Puglia	11.097	12.457	113.124	66.438	135.166	25.054	47.147	282.995	113.599	86.992	76.202
Basilicata	3.377	1.283	17.089	5.827	13.622	2.994	6.457	47.783	16.994	9.824	7.945
Calabria	5.831	3.424	43.125	15.579	43.025	7.464	20.974	92.736	58.808	30.390	14.237
Sicilia	9.324	10.382	189.235	32.712	110.521	15.138	42.853	137.543	110.817	58.657	29.466
Sardegna	4.840	5.387	43.410	19.292	40.944	6.782	19.330	48.863	36.936	22.594	8.236
TOTALE	1.217.540	446.183	3.682.535	2.593.175	3.227.687	709.689	1.999.771	5.066.350	3.065.571	1.096.924	1.204.899
ITALIA	1.217.540	446.183	3.685.020	2.593.175	3.227.687	709.689	2.000.255	5.119.020	3.590.831	1.096.924	1.219.257

28-29 settembre 2011 Statistica sociale

ARRIVI	ITALIA									
	Lazio	Abruzzo	Molise	Campania	Puglia	Basilicata	Calabria	Sicilia	Sardegna	Italia
Piemonte	285.365	35.962	7.528	104.891	52.419	10.218	28.659	69.765	60.121	3.361.727
Valle d'Aosta	8.971	895	132	8.668	3.665	151	1.565	2.623	3.235	106.572
Lombardia	669.468	125.625	13.154	216.765	137.834	21.800	65.053	217.624	202.794	7.724.605
Trentino-A. Adige	31.978	11.898	995	36.999	16.579	1.734	8.949	13.253	15.409	678.403
Veneto	227.940	42.847	6.012	104.546	59.028	7.921	23.711	73.889	63.460	2.847.433
Friuli-V. Giulia	49.747	7.947	1.255	30.395	17.137	2.207	7.272	19.065	13.155	720.882
Liguria	113.011	11.213	1.641	66.260	22.808	3.117	13.340	31.407	28.293	1.030.978
Emilia-Romagna	276.719	60.884	9.248	133.149	71.425	15.321	31.582	73.797	62.945	3.034.614
Toscana	219.139	34.322	5.692	136.849	53.788	8.933	30.133	77.251	53.093	2.271.033
Umbria	41.009	24.047	2.430	65.959	21.740	3.105	8.885	15.817	12.111	619.297
Marche	62.258	36.993	7.372	55.810	34.692	6.318	14.112	24.735	16.045	856.419
Lazio	362.318	213.537	24.081	310.274	171.671	32.036	109.219	199.165	125.353	3.813.268
Abruzzo	85.902	87.739	11.525	39.141	26.109	5.919	11.014	19.651	9.722	669.502
Molise	31.660	14.509	7.936	23.501	10.662	1.851	4.499	5.861	2.439	199.660
Campania	364.816	117.822	22.739	775.337	116.371	41.356	120.367	136.728	39.322	2.964.838
Puglia	418.908	99.620	19.815	134.421	208.380	60.197	88.092	88.739	16.796	2.105.239
Basilicata	39.251	8.968	1.862	48.034	17.188	26.866	13.677	12.835	2.810	304.686
Calabria	220.758	10.946	2.821	69.534	50.355	13.401	109.307	87.739	5.696	906.150
Sicilia	296.440	22.878	5.341	90.594	113.284	15.286	116.679	870.498	22.023	2.299.671
Sardegna	63.967	5.533	667	21.449	14.077	1.367	3.128	21.860	221.201	609.863
TOTALE	3.869.625	974.185	152.246	2.472.576	1.219.212	279.104	809.243	2.062.302	976.023	37.124.840
ITALIA	3.890.451	974.185	152.246	2.472.576	1.219.212	279.104	809.243	2.231.132	1.029.172	37.962.902

28-29 settembre 2011 Statistica sociale

- Anche in questo caso SONO PRESENTI gli elementi sulla diagonale principale: si tratta, nel caso specifico, di buone stime del TURISMO "DOMESTICO" INTRA-REGIONALE

28-29 settembre 2011

Statistica sociale

61

MATRICI DI DISTANZE (O DI DISSOMIGLIANZE)

- In questo tipo di matrici sono rappresentate le distanze tra gli OGGETTI, che sono presenti sia sulle righe che sulle colonne
- Un tipico esempio è la MATRICE DI DISTANZE CHILOMETRICHE, che vedete riprodotta in tutti gli atlanti stradali; ne mostriamo qui un sottoinsieme:

28-29 settembre 2011

Statistica sociale

62

	Genova	Torino	Como	Pavia	Brescia	Bologna
Genova	0	170	185	125	230	295
Torino	170	0	165	155	225	330
Como	185	165	0	85	110	260
Pavia	125	155	85	0	135	200
Brescia	230	225	110	135	0	190
Bologna	295	330	260	200	190	0

28-29 settembre 2011 Statistica sociale 63

PROPRIETÀ DI QUESTO TIPO DI MATRICI

- sono QUADRATE, cioè di dimensioni $k \times k$;
- SONO SEMPRE SIMMETRICHE;
- gli elementi sulla diagonale principale sono TUTTI = 0;
- NON AMMETTONO né TOTALI DI RIGA né TOTALI DI COLONNA.
- I dati che costituiscono questo tipo di matrici sono sempre del tipo
Distanza(A, B)
Dove A e B sono due generici oggetti compresi nella matrice.
- La **distanza euclidea** tra due punti, P e Q, nello spazio k -dimensionale:

$$d_{(P,Q)} = \sqrt{\sum_{j=1}^k \left(x_{jP} - x_{jQ} \right)^2}$$

- è solo un caso particolare di DISTANZA tra due "oggetti"

28-29 settembre 2011 Statistica sociale 64

TABELLE DI CONTINGENZA (O TABELLE A DOPPIA ENTRATA)

- Se si dispone di due variabili (o caratteri) su scala nominale, A e B, le frequenze assolute, conteggiate **SIMULTANEAMENTE** per i due caratteri, si possono organizzare in una tabella di questo tipo, che è detta **TABELLA DI CONTINGENZA**:

28-29 settembre 2011

Statistica sociale

65

	B₁	B₂	...	B_j	...	B_k	
A₁	n_{11}	N_{12}	...	n_{1j}		n_{1k}	$n_{1.}$
A₂	n_{21}	N_{22}	...	n_{2j}		n_{2k}	$n_{2.}$
...							
...							
A_i	n_{i1}	n_{i2}	...	n_{ij}		n_{ik}	$n_{i.}$
...							
...							
...							
A_h	n_{h1}	N_{h2}	...	n_{hj}	...	n_{hk}	$n_{h.}$
	$n_{.1}$	$n_{.2}$...	$n_{.j}$...	$n_{.k}$	N

28-29 settembre 2011

Statistica sociale

66

ESEMPIO

- Abbiamo valutato, in un **campione** di studenti iscritti al III anno di **quattro licei scientifici**, se esista (o meno) una relazione tra il **rendimento scolastico** (espresso dal voto medio di fine anno in tutte le materie) e il **livello di istruzione del capofamiglia**.
- La ricerca, svoltasi su un campione di **180** studenti, ha dato i seguenti risultati (espressi appunto da una tabella di contingenza):

28-29 settembre 2011

Statistica sociale

67

	Lic.media o inferiore	Diploma	Laurea o sup.	TOTALE
Rendimento scarso- insufficiente (3-4-5)	16	12	6	34
Rendimento sufficiente (6-7)	30	55	13	98
Rendimento buono-ottimo (8- 9-10)	3	23	22	48
TOTALE	49	90	41	180

28-29 settembre 2011

Statistica sociale

68

**La stessa tabella si può trasformare in valori percentuali,
NEL SENSO DELLE COLONNE:**

	Lic.media o inferiore	Diploma	Laurea o sup.	TOTALE
Rendimento scarso-insufficiente (3-4-5)	32,7%	13,3%	14,6%	18,9%
Rendimento sufficiente (6-7)	61,2%	61,1%	31,7%	54,4%
Rendimento buono-ottimo (8-9-10)	6,1%	25,6%	53,7%	26,7%
TOTALE	100,0%	100,0%	100,0%	100,0%

Oppure NEL SENSO DELLE RIGHE:

	Lic.media o inferiore	Diploma	Laurea o sup.	TOTALE
Rendimento scarso-insufficiente (3-4-5)	47,1%	35,3%	17,6%	100,0%
Rendimento sufficiente (6-7)	30,6%	56,1%	13,3%	100,0%
Rendimento buono-ottimo (8-9-10)	6,3%	47,9%	45,8%	100,0%
TOTALE	27,2%	50,0%	22,8%	100,0%

28-29 settembre 2011

Statistica sociale

69

Oppure, ancora, RISPETTO AL TOTALE GENERALE:

	Lic.media o inferiore	Diploma	Laurea o sup.	TOTALE
Rendimento scarso-insufficiente (3-4-5)	8,9%	6,7%	3,3%	18,9%
Rendimento sufficiente (6-7)	16,7%	30,6%	7,2%	54,4%
Rendimento buono-ottimo (8-9)	1,7%	12,8%	12,2%	26,7%
TOTALE	27,2%	50,0%	22,8%	100,0%

28-29 settembre 2011

Statistica sociale

70

Grazie per l'attenzione

28-29 settembre 2011

Statistica sociale

71